

# 正倉院文書データベースへのトピックマップ 応用による 奈良時代知識情報構築の試み

後藤 真

内藤 求

花園大学 文学部 (株)ナレッジ・シナジー

正倉院文書データベース (SOMODA) が公開され、おおむね 5 年が経過した。SOMODA は、正倉院文書の復原過程の方法論を情報学的に表現することを目指すと同時に、正倉院文書の総合的な知識情報のデータベースとして作成されたものである。情報技術の進展とともに、SOMODA にも新たな課題が生じるとともに、正倉院文書研究の進展のためにもより高度な研究データベースとしての発展が求められている。そこで、正倉院文書データベースのデータを高度利用し、さらに、関連する知識情報を加え、トピックマップを作成した。このトピックマップの活用により、正倉院文書に関連するデータの、高次の活用が可能となると同時に、新たな知見への可能性も期待できる。

Makoto GOTO  
Faculty of Literature  
Hanazono University

Motomu NAITO  
Knowledge Synergy Inc.

## Experiment of the Nara era knowledge information construction by the topic map apply to Shoso-in document database

Shoso-in Monjo Database (SOMODA) was shown, and about five years passed. SOMODA aimed at expressing the methodology of the reconstruction process of the Shoso-in Monjo for the information science. Additionally, SOMODA was made as a database of the general knowledge information of the Shoso-in Monjo. With progress of the information technology, a new problem occurred in SOMODA. And the development as the higher study database is pursued in SOMODA for the progress of the Shoso-in Monjo study. Therefore, I apply to Topic Maps for the data of the SOMODA. Furthermore, I added associated knowledge information and made a topic map. The good use of this topic map enables the highly advanced use of data related to Shoso-in Monjo. In addition, the possibility of the discovery of new knowledge is also expected.

### 1 . 正倉院文書データベースの課題

著者の一人は、以前、正倉院文書データベース (SOMODA) を作成し、その成果については、人文科学とコンピュータシンポジウムでも継続して発表を重ねてきた[1][2]。SOMODA は 2007 年 3 月に公開され、正倉院文書の復原過程を再現するためのツールとして、利用されてきた[3]。また、正倉院文書に関するデジタル化も、各研究機関で行われるようになってきている。たとえば、東京大学史料編纂所では、正倉院文書の基礎データの提供に関する動きが進みつつあり[4]、国立歴史民俗博物館でも正倉院文書データの高度利用の研

究が進むなど[5]、正倉院文書のデジタル利用がこれまでも増して進みつつある。

一方で、SOMODA にも、いくつかの課題が指摘されるにいたっている。たとえば、画像表示の不具合や、システムの複雑さにもとづく表示の速度の問題などを指摘されていることもあった[6]。また、正倉院文書の複雑さから扱いが難しいという問題点などもあり、今後のブラッシュアップが課題となっている。

### 2 . SOMODA の新たな形

そこで、SOMODA が、今後、どのような形に進むべきか、著者は一つの見通しを示した[7]。それは、以下のとおりである。

- (1) 現在の SOMODA は、正倉院文書研究資料の構造をそのまま採用しているが、それは、正倉院文書の現状とも原状とも異なる。そのため、まずは、データの基礎構造を正倉院文書の現状に即したものにすべきである。
- (2) 検索・情報発見をより容易にするしくみを検討する。
- (3) 正倉院文書に関するデジタル化の進展の状況にかんがみ、多くのデータベースとの協業を可能にするべく、可能な限り Open Data 化を進める。
- (4) 知識情報にもとづく、効果的な情報発見が必要となる。また、史料と研究で得た知識を結びつける必要性がある。

これらを実現するための一つの手法として、オントロジ技術が有用であると、先述の論文では述べた。そして、多くの歴史資料と、歴史の知識情報、その典拠となる論文などについて、それぞれを有機的に結びつけることで、正倉院文書という一つの資料世界に関する総合的なデータベースが可能であるとの指摘を行った。データベースによって正倉院文書に関わる研究全体を俯瞰し、かつ、新知見を得るための重大なヒントを提供し、さらに、正倉院文書に関わる研究の方法論を明示するためである。

そこで、本研究では、そのオントロジ技術として、ISO 標準でもあるトピックマップを採用した。そして SOMODA のデータに加え、新たな研究知識情報をトピックマップの中に入れ、新たな可能性を模索した。以降、本研究について報告する。

### 3 . トピックマップの特徴と優位点

データベースで採用したトピックマップは以下のような特徴を持っている。

トピックマップは、ISO/IEC JTC1 SC34 で策定された ISO 標準である。その標準の中で、以下のように記述されている。"トピックマップは、知識を記号化し、この記号化された知識を関係がある情報資源に結び付ける技術である。トピックマップは、論議の主題 (subjects) を表現するトピック (topics)、主題間の関係を表現する関連 (associations) 及び主題と主題に関連する情報資

源を結び付ける出現 (occurrences) によって組織化することができる[8]。

また、トピック、関連、出現等は、型を持つことができ、それらの型の構造をトピックマップオントロジという。なお、本論文では、このトピックマップオントロジをオントロジと呼称する。

また、トピックマップにおいては、トピックに主題を同定/識別するための主題識別子 (Subject Identifier) を割当てることができる。主題識別子として、IRI (Internationalized Resource Identifier) を利用する。主題識別子を公開したものを公開主題識別子 (PSI: Published Subject Identifier) と呼ぶ。PSI を利用することにより、現在の Web では解決が困難な Synonym (同意語)、Homonym (同音異義語)、Polysemy (多義性) などの問題に煩わされることなく、ユニークな IRI による正確な主題の識別/同定が可能になる。

このトピックマップの特徴により、技術的には以下のメリットが期待できる。

(1) デジタル・アーカイブの構築への適合  
一般的に、デジタル・アーカイブは、管理対象となる情報オブジェクトとそのメタデータ(タイトル、作者、主題、概要、作成日、場所、分類など)から構成される。メタデータは、表構造で表現される場合が多い。しかし、主題、場所、分類などは、ネットワークや階層のデータ構造を持つため、表構造では、それを表現するには困難な場合がある。一方、トピック、関連、出現で構成されるトピックマップのデータ構造は、本来、デジタル化の対象となる多様な文化資源が持つ複雑なデータ構造の表現によく適合すると考えられる。

(2) 多様な索引(アクセスパス)の設定  
文化資源を構成する情報オブジェクト群は、多くの場合、多視点、多層的な知識空間を構成する。従って、画一的、固定的なものでなく、要求に合わせて多様なアクセスパスが柔軟に設定できれば、デジタル・アーカイブの有用性を高めることができる。人文科学研究者の要求に基づいて、視点、知識を体系的に整理し、それをもとにアクセスパスが設定できることが理想である。もっとも、一から体系を構築するのは大変な労力を必要とするため、自ら構築する代わりに、タクソノミ、シソ

ーラスなど既存の知識体系を利用することも選択肢の一つとして与えられている。

### (3) 問合せの優位性

トピックマップ化され設定されたアクセスパスを航行し、目的とする主題についての情報や情報オブジェクトを取得することも可能である。

### (4) 他のアーカイブとの連携

分散して存在しているトピックマップ Web アプリケーション間で、トピックマップのフラグメントの問合せ、交換等をするためのプロトコルとして TMRAP (Topic Maps Remote Access Protocol) が用意されている [9]。TMRAP は、Ontopia (Ontopia 社によって開発され、現在は、Ontopia project によって管理、開発、保守が行われているオープンソースのトピックマップ開発・運用ツール) に含まれている Web サービスインターフェースで、HTTP と SOAP に対応している。これらを用いることで、たとえ異なるトピックマップ上にデータが存在していても、同一の主題に対して、同じ主題識別子を利用していけば、TMRAP を利用して、その主題についての情報を取得することができる。すなわち、デジタル・アーカイブをトピックマップとして構築しておけば、その間の連携が可能になる。

### (5) アーカイブを核とした知識の集積

トピックマップは、PSI のおかげで、意味に基づいて主題の識別が可能であり、主題間の関係にそって、主題間をナビゲートすることができる。そして、主題に関係する情報オブジェクトへのリンクをもつため、あたかもサイバースペース上の意味的なインデックスのような役割を果たすことが可能である。さまざまな領域、文脈に合わせ、同一の情報オブジェクト群に対しても複数の視点からセマンティック・インデックスを作成できる。さらに、それらのインデックスに、関連知識・コメント・評価・情報オブジェクト等への指示も随時追加可能にすることで、デジタル・アーカイブ内部の情報オブジェクトに加え、デジタル・アーカイブとは独立して存在するものを統合的・横断的に利用することが可能になる。つまり、あるデジタル・アーカイブを中核に、関連する知識の集積が可能な環境が構築できる。それにより、デジ

タル・アーカイブの有用性がさらに高まるものと考えられる。

これらのメリットは、正倉院文書の新たな形でのデジタル化に極めて有益であると考えられる。そこで、実際に正倉院文書トピックマップを作成した。次章でその説明を行う。

## 4. 作成した正倉院文書トピックマップ

筆者らは、以下のように正倉院文書のトピックマップを作成した。作成したオントロジ図は、図 1 のとおりである。

### (1) 正倉院文書の現状のテキストの構造化と永続的識別子の付与

まず、正倉院文書の歴史的経緯と構造を踏まえ、正倉院文書の現状と、奈良時代の「復原」状況である原状の構造の関係を記した。正倉院文書の現状は、正集～続々修といった所属 帙 巻 断簡 (奈良時代の原状を残している紙の最小単位) 紙という階層構造を持っている。その階層構造の表現を実現させた。また、原状の構造を帳簿 (復原された文書) 断簡 紙という構造をもって記した。

そして、紙と断簡に PSI を付した。これにより断簡 ID・紙 ID を URI で参照することで、正倉院文書の史料情報を、別のデータベースで読み込み、高度利用が可能な形式とし、正倉院文書デジタルデータのさらに高度な利用を可能とした。また、他の研究論文やデータベースから、個別の断簡への永続的ポイントを示すことを可能にすることで、正倉院文書トピックマップが研究資源として安定的に利用できるような仕組みを可能とした。

ただし、正倉院文書の断簡は、今後の調査状況によって変更される可能性があるため、永続的な識別子をどのように詳細に定義するかは今後の課題である。

この資料の現状構造を基本とした典拠情報群が、正倉院文書のトピックマップ全体の基礎情報となる。

(2) 正倉院文書に関する研究知識情報の付与  
基礎情報を付加した次のステップとして、正倉院文書に関連する知識情報を整理し、主題化し、史料と結びつけた。

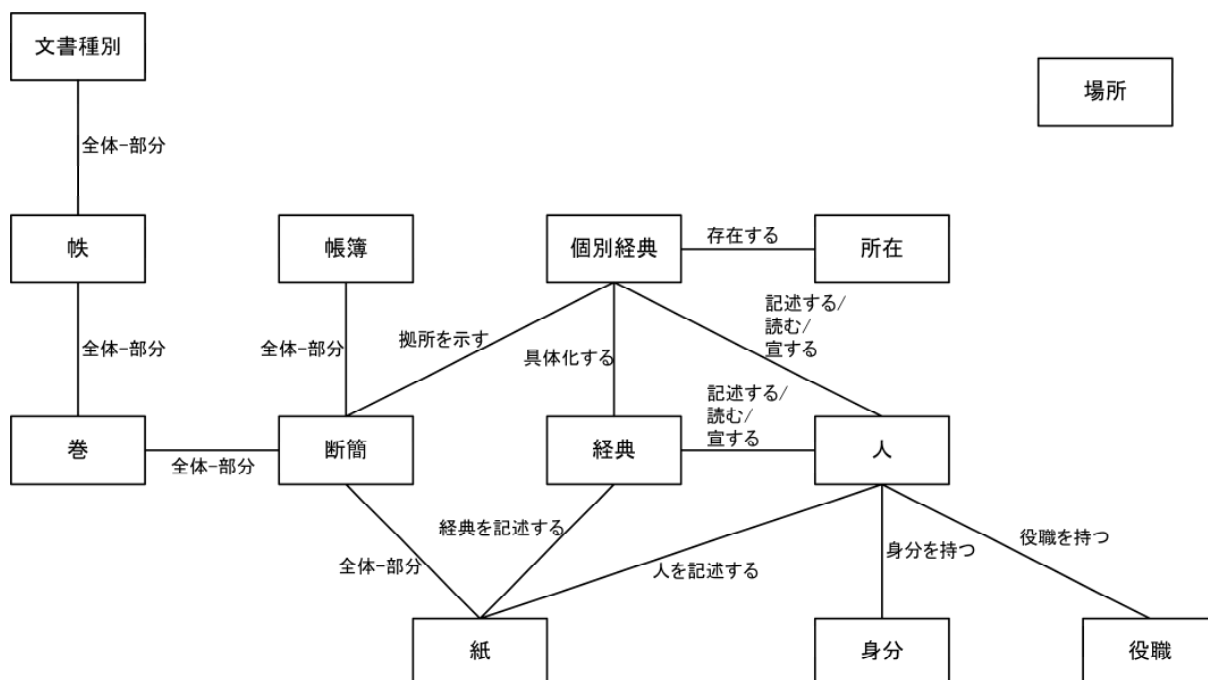


図1 作成した正倉院文書トピックマップのオントロジ図  
 四角はトピック型を表し、四角の中の文字列はトピック型の名前を表す。  
 線は関連型を表し、線の中央付近の文字列は関連型の名前を表す

正倉院文書の中心は、奈良時代の写経機関（造東大寺司写経所）の帳簿である。また、近年の正倉院文書研究の中心は、その写経所に関するものとなっている[10]。そのような研究状況に対応すべく、写経所に関連する情報と人物に関連する情報について記述することを、知識情報付与に関する最初のステップとした。

奈良時代の人名に関する情報のうち、正倉院文書に出てくるものを既存の辞書類から抽出した。そのうえで、既存の日本史の研究成果に基づき[11]、筆者らが人名に対し、身分と、役職、そして居所について関係づけを行った。

そして、これらの人名が出てくる正倉院文書の基礎構造である「紙」と関連づけた。これにより、正倉院文書に記述された人についての典拠情報が明示される。単純に人名を文字列で検索するだけでなく、同一人物で別名や、略称などを同時に発見できる。また、同一役職の人物を同時に発見し、原史料にあたることのできるようになり、セマンティックな検索を実現できるようになる。

知識情報の付加の第2ステップとして、奈良時代に存在したことが史料上明確な經典に関する情報を既存の辞書類から抽出した。經典に関しては、それぞれに「作者」を付し、また、判明しているものについては、大正新脩大蔵經の番号を付した。

しかし、經典名だけで正倉院文書の中に出てくる經典を記述することは困難であった。なぜなら、コンテンツとして記載される經典名と、史料上で出てくる実物とは扱いが異なるためである。そこで、經典名と、実際に史料中に出てくる「モノ」としての經典を関連付けた。これは書名と実体としての書籍の関係として理解できる。

## 5. 正倉院文書トピックマップの有用性

正倉院文書トピックマップ（第1版）では、具体的に以下のような歴史学に対する有用性が期待されている。

### (1) 正倉院文書の人名と經典名の結びつけ

写経所において、ある人物が、特異な經典に関わっていたことがあるのかなど、通常の検索モデルだけではわからないセマンティックな情報の発

見が期待できる。

たとえば図 2 は、作成したトピックマップをもとに奈良時代のある時点での經典注釈書である疏の保持者とその所在地の関係を Visualize したものである。これにより、ある僧が持っていた注釈書群がどのようなものか、また、その注釈書の作者に偏りがあるのかなどがわかる。

また、史料上判明した僧の居所についてもトピックマップに入っているので、これにより、寺院-經典-僧の関係性を、經典の所持という観点から明らかにできる可能性がある。

さらに、時間情報を入れておくことで、その經典類が時間の変遷とともに保持者が変わるのか、所在場所が変わるのかなどの情報を得ることができる。上記の情報や研究をもととした古代国家の仏教編成などについて、中林隆之氏が継続的に研究を行っており[12][13]、その研究の支援へとつながることも期待できる。

(2) 帳簿と經典・人名の結びつけ

正倉院文書は、その歴史的経緯から、大変複雑な構造を持っており、それが、正倉院文書の全体

像を見えにくくし、研究を困難にしてきた背景がある。それを解決したのが SOMODA であった。しかし、SOMODA は、一つの帳簿を復原し、その帳簿の状況を概観するのは優位であるものの、特定のキーワードとそれに関係する帳簿を横断的に複数見るなどには向かない点があった。正倉院文書トピックマップでは、それらの問題点を解決することが期待できる。

特定の經典が、ある特殊な帳簿に出てくる可能性があるのか、また、ある人名の集団が、特定の帳簿に出てくるなどの傾向があるのかの分析は、奈良時代の実態に迫ると同時に、当時の帳簿の記述形式や手法といった、正倉院文書の史料論に結び付けられる。

たとえば、造東大寺司写経所には、東堂・西堂、もしくは北堂・南堂といった二つの堂舎によって作業が行われていたことが指摘されている。しかし、その詳細は必ずしもわかっているとはいえない。帳簿ごとに人名集団の傾向が判明すれば、このような実態の研究に迫ることが可能になると考えられる。

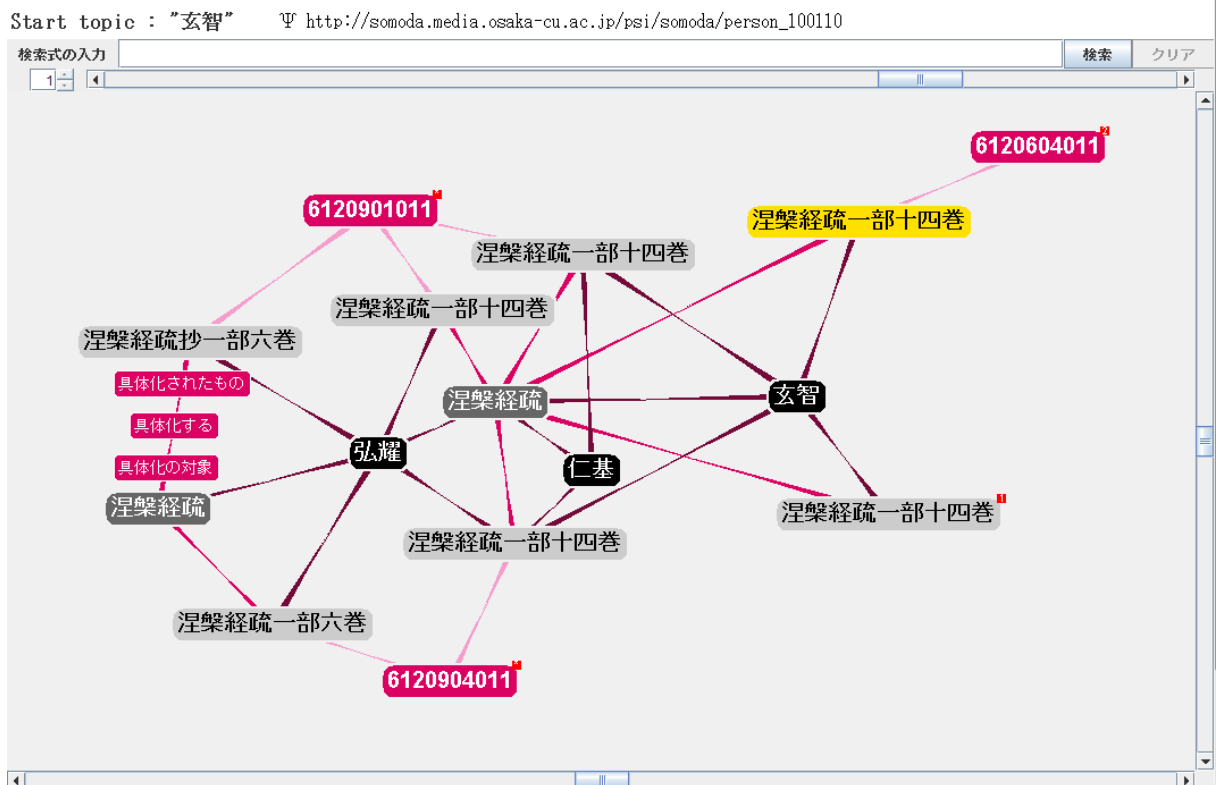


図 2 僧名を基礎に經典と資料 ID の関連付けを Visualize したものの

### (3) 人名同士の結びつけ

また、人名同士や、役職同士である特定のつながりがあるのかなどを分析することが可能となる。これにより、今までは漠として指摘されていた、正倉院文書の記述の中心である造東大寺司写経所の人間関係などがわかる可能性がある。写経に従事する時期の差や、作業状況なども、人名をセマンティックに分析することで可能となる部分もあるであろう。また、栄原永遠男氏が行った、月借錢解に関する、人間関係の分析なども、より精密に検討できる可能性を秘めている。

このように、正倉院文書トピックマップ第1版として、経典と人名に関する情報を入れただけでも相当の人文科学研究への貢献が期待できる。

将来的な正倉院文書トピックマップでは、これ以外にも、多くの研究知識情報と正倉院文書の史料そのものを有機的に結び付け、正倉院文書研究に関する知識情報の到達点を情報学的に提示すると同時に、次のステップへと役立てることを目指したい。

## 6 . 課題

本研究で作成しているデータベースは緒についたばかりであり、以下のような課題がある。

### (1) 複雑な時間遷移の処理

正倉院文書は奈良時代という比較的短い期間の文書群ではあるが、時間的な変化はあり、かつ、その時間的な変化のなかでオントロジの構造が変わる。

たとえば、人名に関しては、役職が変わることもあり、内部での昇進なども行われている。経典に関して、時間的な変遷とともに場所を移動するのが一般的である。

この変化をどのように処理していくかは今後の課題である。また、正倉院文書の記述内容そのものは、奈良時代のものだが、その後の複雑な史料構造の変化まで含める場合には、江戸時代や明治時代まで考慮の対象とする必要がある。

オントロジにおける、適切なドメインの設定の重要性はいくつか指摘されているが、本トピック

マップにも、その課題はあてはまるものと考ええる。正倉院文書という、ある種史料的に「閉じた」世界を記述することを試みたとしても、これほどの広がりがある。

このような状況にどのように対応するのか、今後の課題といえよう。

### (2) 適切なトピックの設定

第一段階では、人名・経典名といった史料に即した情報を入力しているが、トピックマップの本来の目的にそうならば、より抽象的な研究知識情報を入力することが求められる。このトピックの設定が正倉院文書研究にどこまで即しているのかを常に考慮しておく必要がある。

トピックマップの構築は、いわば、正倉院文書研究の、現時点での到達点を、デジタルで記述することを目指している。研究史を適切に描くための工夫と、研究実態に即したトピックの設定が重要になる。

次の段階として、正倉院文書の研究書に付された索引を二次利用させていただき許諾を得ており、それを応用することを次の構想としている。

### (3) PSI の課題

SOMODA のときからの継続的課題ではあるのだが、正倉院文書の断簡は現在も原本調査が続いており、いまだに、確定した情報を持っているわけではない。SOMODA は、仮想的・実験的にその断簡を設定し、復原することを目指したデータベースであった。それゆえ、現在でも確定していない断簡を処理するのは、いわば「当然」であり、問題はなかった。

しかし、正倉院文書トピックマップは、帳簿の構造をオントロジで示すため、また、研究情報の横断的提供のため、断簡を永続的識別子として定義する必要がある。この段階で、再度断簡の未確定問題にあたることになってしまう。

断簡の情報を、正倉院文書研究の側で仮に共通で呼称する何らかの工夫が必要になるであろう。この点は、正倉院文書研究者と、データベース構築側の適切な対話が必要となる。

## 7 . 展望

今後は以下のような展望を描いている。

### (1) オントロジの拡張

前章の課題でも述べたが、研究情報をより効果的に表現することを目指している。役職を構造化し、人物の関係を効果的に表現する、経典の教義関係を整理し、経典の構造を描くなどを第一段階として、より抽象的な研究知識情報の構築を目指す。

### (2) 他史料とのリンク

正倉院文書の時代は、奈良時代である。この同時代史料として、『続日本紀』や木簡などがある。これらの史料との関係を構築することで、奈良時代の全体像をデジタルで記述することも可能となるであろう。

『続日本紀』については、現在、時空間情報を持つ記述を抽出し、Hutime（京都大学地域研究統合情報センターが中心となっている HGIS 研究会によって開発された時空間分析ツール）への適用を試みるなどの動きを進めると同時に、TEI への適用実験を試みている[13]。この試みは、『続日本紀』をテキストとしてだけでなく、多様なかたちで、かつ他のデジタル・アーカイブとの横断的に利用できることを目指している。正倉院文書トピックマップとの連携によって、奈良時代の全体像を、より具体的に描くことが可能となるであろう。

また、正倉院文書研究の側からは、とかく「閉じた」世界として批判されやすい正倉院文書研究を奈良時代全体の研究と接続させることで、研究のステップをもう一段階先に進めることができると考えられる。

さらに、将来的には、木簡データベースなどとの接続が可能となれば、データベースによる奈良時代史研究への道筋が描ける可能性もあるであろう。

また、正倉院文書に多く出てくる経典に関するデータベースとの接続も目指したい。正倉院文書研究は、日本史学と国文学が多く中心を占めていて、仏教学の立場からの研究は決して多いとはいえない。これは、正倉院文書の複雑さもさることながら、正倉院文書の経典情報が、学的につながっていない部分があるのではないかと考える。そこで、経典に関するデータベースと接続すること

で、この課題が解決できるのではないだろうか。たとえば、正倉院文書に出てくる経典名から、実際の経典テキストへのリンクを作成する、もしくは、経典データベースから、該当する正倉院文書の帳簿・断簡へリンクを作成することで仏教研究者の正倉院文書へのルートが確保できる。

### (3) 他の正倉院文書のリソースへのリンク

本稿執筆現在、Web 上における正倉院文書のデジタルデータの提供は SOMODA のみである。しかし、1 章で述べたとおり、いくつかの正倉院文書に関連する機関において、正倉院文書の高度情報化に関する検討が行われている。今後、これらの成果がリソースとして公開されるならば、それらのリソースとのリンクを作成することで、正倉院文書の総体的なデジタルデータの提供が可能となると考えられる。

正倉院文書トピックマップは、情報技術や Web の研究動向にあわせた新しい人文情報研究の形を目指したものであるといえる。

## 8 . 人文情報学との接続に向けて

正倉院文書トピックマップは、正倉院文書の研究の方法論をデジタルで記述することを試みた SOMODA の発展形として位置づけられる。SOMODA は、正倉院文書の復原過程を中心に記述したが、正倉院文書トピックマップは、そこからさらに研究資料の知識情報もあわせて記述することを目指した。

正倉院文書の研究課題の解決や、総体的な研究、そして、研究資源の提供という意味において、人文情報学的な貢献が可能となるのではないかと考える。

また、奈良時代史を研究する際の、情報のポータルであり、ハブとして正倉院文書トピックマップが位置づけられることを目指している。日本史における史料デジタル・アーカイブはいまだに散在しているが、それを時代ごとにまとめて「見る」手法の一提案としても正倉院文書トピックマップを位置づけることができる。史料そのもののデータベースは各館で持ちつつも、ポータルを外部に作る手法である。長期的には、奈良時代の知

識情報全体を覆う辞書として、発展することを最終形としている。

日本史学は、属性の異なる資料をメタ、かつ横断的に見て分析を行うという研究手法の特徴がある。その研究手法の特性により、データベースの効果的な応用が隣接他分野に比して進みにくいという側面がある。正倉院文書トピックマップがその課題を効果的に解決する一つの手法となれば幸いである。

## 参考文献

- [1] 後藤真 柴山守「正倉院文書研究資料のXML/XSLTによる記述と統合」『人文科学とコンピュータシンポジウム 2002』pp209-216、2002年
- [2] 後藤真 柴山守「正倉院文書データベースにおける検索と人文科学的研究成果との連携」『人文科学とコンピュータシンポジウム 2005』pp209-216、2005年
- [3] <http://somoda.media.osaka-cu.ac.jp>, 正倉院文書データベース (SOMODA)
- [4] 東京大学史料編纂所特定共同研究「古代史料の研究資源化」「正倉院文書に関する史料学情報の研究資源化連携」
- [5] 人間文化研究機構連携研究「人間文化資源」の総合的研究」国立歴史民俗博物館「正倉院文書の高度情報化研究」
- [6] 小口雅史「日本古代史研究のためのオンライン・データベース」『日本歴史』740号、pp9-15、2010年
- [7] 後藤真「デジタル技術による歴史情報の可視化の試み 奈良時代知識データベース構築試論」栄原永遠男編『日本古代の王権と社会』、pp343 - 360、2010年
- [8] 内藤求編著『トピックマップ入門』東京電機大学出版局、2006年
- [9] <http://www.ontopia.net/topicmaps/tmrap.html>  
TMRAP : Topic Maps Remote Access Protocol, Ontopia.
- [10] 栄原永遠男『奈良時代の写経と内裏』塙書房、2000年
- [11] 山下有美『正倉院文書と写経所の研究』吉川弘文館、1999年
- [12] 中林隆之「東アジア<政治-宗教>世界の形成と日本古代国家」『歴史学研究』885号、2011年
- [13] 科学研究費基盤研究(C)「経典目録よりみた古代国家の宗教編成策に関する多面的研究」研究代表者：中林隆之

- [14] 後藤真「HuTime/Mapの日本史研究への応用の試み 続日本紀を題材に」HGIS研究会編『HuTime/Mapを使った研究事例と将来展望』報告書、2012年

本研究はJSPS科研費 22520659, 24242025の助成を受けたものである。